Linear algebra and analysis recalls

Lectures for PHD course on Numerical optimization

Enrico Bertolazzi

DIMS – Universitá di Trento

November 21 – December 14, 2011

Outline

Outline

- We always work with finite dimensional Euclidean vector spaces $\mathbb{R}^n$, the natural number n denote the dimension of the space.
- Elements $\boldsymbol{v} \in \mathbb{R}^n$ will be referred to as vectors, and we think them as composed of n real numbers stacked on top of each other, i.e.,

$$\boldsymbol{v} = \left(v_1, v_2, \dots, v_n\right)^T = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}$$

v_k being real numbers, and T denotes the transpose operator.

Basic operation

Basic operations defined for two vectors $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^n$, and an arbitrary scalar $\alpha \in \mathbb{R}$

$$\boldsymbol{a} = \left(a_1, a_2, \dots, a_n\right)^T \qquad \boldsymbol{b} = \left(b_1, b_2, \dots, b_n\right)^T$$

are:

1. addition: $\boldsymbol{a} + \boldsymbol{b} = \left(a_1 + b_1, \dots, a_n + b_n\right)^T \in \mathbb{R}^n$;
2. multiplication by a scalar: $\alpha \boldsymbol{a} = \left(\alpha a_1, \dots, \alpha a_n\right)^T \in \mathbb{R}^n$;
3. scalar product between two vectors: $(\boldsymbol{a}, \boldsymbol{b}) = \boldsymbol{a}^T \boldsymbol{b} = \sum_{k=1}^n a_i b_i \in \mathbb{R}$.
4. A linear subspace $L \subset \mathbb{R}^n$ is a set with the two properties:
   1. for every $\boldsymbol{a}, \boldsymbol{b} \in L$ it holds that $\boldsymbol{a} + \boldsymbol{b} \in L$;
   2. and for every $\alpha \in \mathbb{R}$, $\boldsymbol{a} \in L$ it holds that $\alpha \boldsymbol{a} \in L$.
5. An affine subspace $A \subset \mathbb{R}^n$ is any set that can be represented as $\boldsymbol{v} + L := \{\boldsymbol{v} + \boldsymbol{x} | \boldsymbol{x} \in L\}$ for some vector $\boldsymbol{v} \in \mathbb{R}^n$ and some linear subspace $L \subset \mathbb{R}^n$.

Norm

- We associate a norm, or length, of a vector $\boldsymbol{v} \in \mathbb{R}^n$ with a scalar product as:

$$\|\boldsymbol{v}\| = \sqrt{(\boldsymbol{v}, \boldsymbol{v})}$$

- The Cauchy–Bunyakowski–Schwarz inequality says that

$$(\boldsymbol{a}, \boldsymbol{b}) \leq \|\boldsymbol{a}\| \, \|\boldsymbol{b}\| \qquad \text{for } \boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^n$$

- we define the angle θ between two vectors via

$$\cos\theta = \frac{(\boldsymbol{a}, \boldsymbol{b})}{\|\boldsymbol{a}\| \, \|\boldsymbol{b}\|}.$$

- We say that $\boldsymbol{a}$ is orthogonal to $\boldsymbol{b}$ if and only if $(\boldsymbol{a}, \boldsymbol{b}) = 0$.
- The only vector orthogonal to itself is $\mathbf{0} = (0, \ldots, 0)^T$; moreover, this is the only vector with zero norm.

Linear and affine dependence

- The scalar product is symmetric and bilinear, i.e., for every $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$, α, β it holds that $(\boldsymbol{a}, \boldsymbol{b}) = (\boldsymbol{b}, \boldsymbol{a})$, and

$$(\alpha \boldsymbol{a} + \beta \boldsymbol{b}, \boldsymbol{c}) = \alpha(\boldsymbol{a}, \boldsymbol{c}) + \beta(\boldsymbol{b}, \boldsymbol{c})$$

- A collection of vectors $(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k)$ is said to be linearly independent if and only if

$$\sum_{i=1}^{k} \alpha_i \boldsymbol{v}_i = \mathbf{0} \qquad \Rightarrow \qquad \alpha_1 = \cdots = \alpha_k = 0.$$

- Similarly, a collection of vectors $(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k)$ is said to be affinely independent if and only if the collection $(\boldsymbol{v}_2 - \boldsymbol{v}_1, \boldsymbol{v}_3 - \boldsymbol{v}_1, \ldots, \boldsymbol{v}_k - \boldsymbol{v}_1)$ is linearly independent.

Basis

- The largest number of linearly independent vectors in $\mathbb{R}^n$ is n;
- n linearly independent vectors from $\mathbb{R}^n$ is referred to as basis.
- The basis $(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n)$ is said to be orthogonal if $(\boldsymbol{v}_i, \boldsymbol{v}_j) = 0$ for all $i \neq j$. If, in addition $\|\boldsymbol{v}_i\| = 1$ for $i = 1, \ldots, n$, the basis is called orthonormal.
- Given the basis $(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n)$ every vector $\boldsymbol{v}$ can be written in a unique way as $\boldsymbol{v} = \sum_{i=1}^n \alpha_i \boldsymbol{v}_i$, and the n-tuple $(\alpha_1, \ldots, \alpha_n)$ will be referred to as coordinates of $\boldsymbol{v}$ in this basis.
- If the basis $(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n)$ is orthonormal, the coordinates α_i are computed as $\alpha_i = (\boldsymbol{v}, \boldsymbol{v}_i)$.
- The space $\mathbb{R}^n$ will be typically equipped with the standard basis $(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n)$ where $\boldsymbol{e}_i = (0, \ldots, 0, 1, 0, \ldots, 0)^T$.
- For every vector $\boldsymbol{v} = (v_1, \ldots, v_n)^T$ we have $(\boldsymbol{v}, \boldsymbol{e}_i) = v_i$ which allows us to identify vectors and their coordinates.

Matrices

- All linear functions from $\mathbb{R}^n$ to $\mathbb{R}^k$ may be described using a linear space of real matrices $\mathbb{R}^{k\times n}$ (i.e., with k row and n columns).
- Given a matrix $\boldsymbol{A} \in \mathbb{R}^{k\times n}$ it will often be convenient to view it as a row of its columns, which are thus vectors in $\mathbb{R}^k$.
- Let $\boldsymbol{A} \in \mathbb{R}^{k\times n}$ have elements A_{ij} we write $\boldsymbol{A} = (\boldsymbol{a}_1, \dots, \boldsymbol{a}_n)$, where $\boldsymbol{a}_i = (A_{1i}, \dots, A_{ki})^T \in \mathbb{R}^k$.
- The addition of two matrices and scalar-matrix multiplication are defined in a straightforward way. For $\boldsymbol{v} = (v_1, \dots, v_n) \in \mathbb{R}^n$ we define

$$\boldsymbol{A}\boldsymbol{v} = \sum_{i=1}^{n} v_i \boldsymbol{a}_i \in \mathbb{R}^k$$

Matrix norm and transpose

- We also define a norm of the matrix $\boldsymbol{A}$ by

$$\|\boldsymbol{A}\| = \max_{\boldsymbol{v}\in\mathbb{R}^n, \|\boldsymbol{v}\|=1} \|\boldsymbol{A}\boldsymbol{v}\|$$

- For a given matrix $\boldsymbol{A} \in \mathbb{R}^{k\times n}$ we define $\boldsymbol{A}^T \in \mathbb{R}^{n\times k}$ with elements $(\boldsymbol{A}^T)_{ij} = A_{ji}$ as matrix transpose
- A more elegant definition: $\boldsymbol{A}^T$ is the unique matrix, satisfying the equality $(\boldsymbol{A}\boldsymbol{v}, \boldsymbol{u}) = (\boldsymbol{v}, \boldsymbol{A}^T\boldsymbol{u})$ for all $\boldsymbol{v} \in \mathbb{R}^n$ and $\boldsymbol{u} \in \mathbb{R}^k$.
- From this definition it should be clear that $\|\boldsymbol{A}\| = \left\|\boldsymbol{A}^T\right\|$ and that $(\boldsymbol{A}^T)^T = \boldsymbol{A}$

Matrix product

- Given two matrices $\boldsymbol{A} \in \mathbb{R}^{k\times n}$ and $\boldsymbol{B} \in \mathbb{R}^{n\times m}$, we define the product matrix product $\boldsymbol{C} = \boldsymbol{AB} \in \mathbb{R}^{k\times m}$ elementwise by

$$C_{ij} = \sum_{\ell=1}^{n} A_{i\ell} B_{\ell j}, \qquad i = 1, \ldots, k \quad j = 1, \ldots, m.$$

- In other words, $\boldsymbol{C} = \boldsymbol{AB}$ iff for all $\boldsymbol{v} \in \mathbb{R}^n$, $\boldsymbol{Cv} = \boldsymbol{A}(\boldsymbol{Bv})$.
- The matrix product is:
  - associative i.e., $\boldsymbol{A}(\boldsymbol{BC}) = (\boldsymbol{AB})\boldsymbol{C}$;
  - not commutative i.e., $\boldsymbol{AB} \neq \boldsymbol{BA}$ in general;

  for matrices of compatible sizes.

Matrix norm and product

- It is easy (and instructive) to check that

$$\|\boldsymbol{AB}\| \leq \|\boldsymbol{A}\| \, \|\boldsymbol{B}\|$$

  and that $(\boldsymbol{AB})^T = \boldsymbol{B}^T \boldsymbol{A}^T$.
- Vectors $\boldsymbol{v} \in \mathbb{R}^n$ can be (and sometimes will be) viewed as matrices $\boldsymbol{v} \in \mathbb{R}^{n \times 1}$.
- Check that this embedding is norm-preserving, i.e., the norm of $\boldsymbol{v}$ viewed as a vector equals the norm of $\boldsymbol{v}$ viewed as a matrix with one column.
- The triangle inequality for vectors and matrices is valid

$$\|\boldsymbol{a} + \boldsymbol{b}\| \leq \|\boldsymbol{a}\| + \|\boldsymbol{b}\| \,, \qquad \|\boldsymbol{A} + \boldsymbol{B}\| \leq \|\boldsymbol{A}\| + \|\boldsymbol{B}\|$$

$$\|\boldsymbol{a} - \boldsymbol{b}\| \geq \|\boldsymbol{a}\| - \|\boldsymbol{b}\| \,, \qquad \|\boldsymbol{A} - \boldsymbol{B}\| \geq \|\boldsymbol{A}\| - \|\boldsymbol{B}\|$$

Matrix inverse

- For a square matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ we can discuss the existence of the unique matrix $\boldsymbol{A}^{-1}$, called the inverse of $\boldsymbol{A}$, verifying $\boldsymbol{A}^{-1}\boldsymbol{A}\boldsymbol{v} = \boldsymbol{v}$ for all $\boldsymbol{v} \in \mathbb{R}^n$.
- If the inverse of a given matrix exists, we call the latter nonsingular. The inverse matrix exists iff
  - the columns of $\boldsymbol{A}$ are linearly independent;
  - the columns of $\boldsymbol{A}^T$ are linearly independent;
  - the system $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{v}$ has a unique solution for every $\boldsymbol{v} \in \mathbb{R}^n$;
  - the system $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$ has $\boldsymbol{x} = \boldsymbol{0}$ as its unique solution.
- From this definition it follows that $\boldsymbol{A}$ is nonsingular iff $\boldsymbol{A}^T$ is nonsingular, and, furthermore, $(\boldsymbol{A}^{-1})^T = (\boldsymbol{A}^T)^{-1}$ and therefore will be denoted simply as $\boldsymbol{A}^{-T}$.
- At last, if $\boldsymbol{A}$ and $\boldsymbol{B}$ are two nonsingular matrices of the same size, then $\boldsymbol{A}\boldsymbol{B}$ is nonsingular and $(\boldsymbol{A}\boldsymbol{B})^{-1} = \boldsymbol{B}^{-1}\boldsymbol{A}^{-1}$.

Eigenvalues and eigenvectors (1/2)

- If for some vector $\boldsymbol{v} \in \mathbb{R}^n$, and some scalar $\alpha \in \mathbb{R}$ it holds that $\boldsymbol{A}\boldsymbol{v} = \alpha \boldsymbol{v}$, we call α an eigenvalue of $\boldsymbol{A}$ and $\boldsymbol{v}$ an eigenvector, corresponding to eigenvalue α.
- Eigenvectors, corresponding to a given eigenvalue, form a linear subspace of $\mathbb{R}^n$; two nonzero eigenvectors, corresponding to two distinct eigenvalues are linearly independent.
- In general, every matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ has n eigenvalues (counted with multiplicity), maybe complex, which are furthermore roots of the characteristic equation $\det(\boldsymbol{A} - \lambda \boldsymbol{I}) = 0$, where $\boldsymbol{I} \in \mathbb{R}^{n \times n}$ is the identity matrix, characterized by the fact that for all $\boldsymbol{v} \in \mathbb{R}^n : \boldsymbol{I}\boldsymbol{v} = \boldsymbol{v}$.

Eigenvalues and eigenvectors (2/2)

- In general we have $\|\boldsymbol{A}\| \geq |\lambda_n|$ where λ_n is the eigenvalue with largest absolute value.
- The matrix $\boldsymbol{A}$ is nonsingular iff none of its eigenvalues are equal to zero, and in this case the eigenvalues of $\boldsymbol{A}^{-1}$ are equal to the reciprocal of the eigenvalues of $\boldsymbol{A}$.
- The eigenvalues of $\boldsymbol{A}^T$ are equal to the eigenvalues of $\boldsymbol{A}$.
- We call $\boldsymbol{A}$ symmetric iff $\boldsymbol{A}^T = \boldsymbol{A}$. All eigenvalues of symmetric matrices are real, and eigenvectors corresponding to distinct eigenvalues are orthogonal.

Outline

Taylor series

A function $f(x)$ has the expansion

$$f(x+h) = f(x) + hf'(x) + \cdots + \frac{h^k}{k!} f^{(k)}(x) + E$$

where the error term E take the forms

$$\begin{aligned} E &= \frac{1}{k!} \int_0^h (h-t)^k f^{(k+1)}(x+t)\,\mathrm{d}t, && && \text{[Peano]} \\ &= \frac{h^{k+1}}{(k+1)!} f^{(k+1)}(x+\eta), && \eta \in (0,h) && \text{[Lagrange]} \\ &= \mathcal{O}(h^{k+1}) \end{aligned}$$

Multi-index notation

Given a *list* of (non negative) integer $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_n)$ called multi-index and a vector $\boldsymbol{z} \in \mathbb{R}^n$ and a function $f : \mathbb{R}^n \mapsto \mathbb{R}$ we define

- $\boldsymbol{\alpha}! = \alpha_1!\,\alpha_2! \cdots \alpha_n!$
- $|\boldsymbol{\alpha}| = \alpha_1 + \alpha_2 + \cdots + \alpha_n$
- $\boldsymbol{z}^{\boldsymbol{\alpha}} = z_1^{\alpha_1} z_2^{\alpha_2} \cdots z_n^{\alpha_n}$
- $\dfrac{\partial f(\boldsymbol{z})}{\partial \boldsymbol{\alpha}} = \dfrac{\partial^{|\boldsymbol{\alpha}|} f(z_1, z_2, \dots, z_n)}{\partial \alpha_1 \partial \alpha_2 \cdots \partial \alpha_n}$

Multivariate Taylor series

A function $f : \mathbb{R}^n \mapsto \mathbb{R}$ has the expansion

$$f(\boldsymbol{x} + \boldsymbol{h}) = \sum_{|\boldsymbol{\alpha}|=0}^{k} \frac{\boldsymbol{h}^{\boldsymbol{\alpha}}}{\boldsymbol{\alpha}!} \frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{\alpha}} + E$$

where the error term E take the forms

$$E = (k+1) \sum_{|\boldsymbol{\alpha}|=k+1} \frac{\boldsymbol{h}^{\boldsymbol{\alpha}}}{\boldsymbol{\alpha}!} \int_0^1 (1-t)^k \frac{\partial f(\boldsymbol{x} + t\boldsymbol{h})}{\partial \boldsymbol{\alpha}} \, \mathrm{d}t$$

$$= \mathcal{O}(\|\boldsymbol{h}\|^{k+1})$$

Multivariate Taylor series, second order special case

A function $f : \mathbb{R}^n \mapsto \mathbb{R}$ has the expansion

$$f(\boldsymbol{x}+\boldsymbol{h}) = f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})\boldsymbol{h} + \frac{1}{2}\boldsymbol{h}^2\nabla^2 f(\boldsymbol{x})\boldsymbol{h} + \mathcal{O}(\|\boldsymbol{h}\|^3)$$

where

$$\nabla f(\boldsymbol{x}) = (\partial_{x_1} f, \partial_{x_2} f, \ldots, \partial_{x_n} f),$$

$$\nabla^2 f(\boldsymbol{x}) = \begin{pmatrix} \partial^{(2)}_{x_1} f & \partial_{x_1}\partial_{x_2} f & \cdots & \partial_{x_1}\partial_{x_n} f \\ \partial_{x_1}\partial_{x_2} f & \partial^{(2)}_{x_2} f & \cdots & \partial_{x_2}\partial_{x_n} f \\ \vdots & & & \vdots \\ \partial_{x_1}\partial_{x_n} f & \partial_{x_2}\partial_{x_n} f & \cdots & \partial^{(2)}_{x_n} f \end{pmatrix}$$

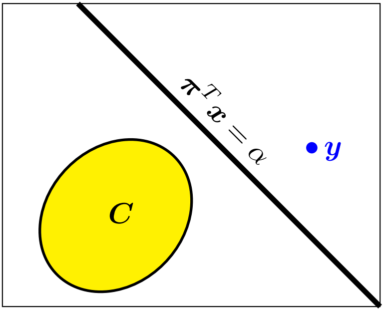
Outline

The Separation Theorem

Theorem (Separation Theorem)

Let be $C \subseteq \mathbb{R}^n$ closed and convex, and $\boldsymbol{y} \notin C$.
Then there exist a real α and a vector $\boldsymbol{\pi} \neq \mathbf{0}$ such that:

1. $\boldsymbol{\pi}^T \boldsymbol{y} > \alpha$;
2. $\boldsymbol{\pi}^T \boldsymbol{x} \leq \alpha$ *for all* $\boldsymbol{x} \in C$.



Proof.

Define the function $\mathsf{f} : \mathbb{R}^n \mapsto \mathbb{R}$ by $\mathsf{f}(\boldsymbol{x}) = \frac{1}{2}\|\boldsymbol{x}-\boldsymbol{y}\|^2$. Now by the Weierstrass Theorem there exists $\boldsymbol{z} \in C$ such that:

$$\mathsf{f}(\boldsymbol{z}) \leq \mathsf{f}(\boldsymbol{x}), \qquad \forall \boldsymbol{x} \in C$$

due to the convexity of C we have $\boldsymbol{z} + t(\boldsymbol{x}-\boldsymbol{z}) \in C$ for all $t \in [0,1]$ and then

$$0 \leq \frac{\mathsf{f}(\boldsymbol{z}+t(\boldsymbol{x}-\boldsymbol{z})) - \mathsf{f}(\boldsymbol{z})}{t},$$

taking the limit $t \to 0$ and noting that $\nabla\mathsf{f}(\boldsymbol{x}) = \boldsymbol{x}-\boldsymbol{y}$ we have

$$0 \leq \nabla\mathsf{f}(\boldsymbol{z})(\boldsymbol{x}-\boldsymbol{z}) = (\boldsymbol{z}-\boldsymbol{y})^T(\boldsymbol{x}-\boldsymbol{z})$$

Now setting $\boldsymbol{\pi} = \boldsymbol{y}-\boldsymbol{z}$ and $\alpha = \boldsymbol{\pi}^T\boldsymbol{z}$ gives the result. □

The Farkas's lemma

Lemma (Farkas's lemma)

Let $\boldsymbol{A} \in \mathbb{R}^{n\times m}$*,* $\boldsymbol{b} \in \mathbb{R}^n$ *and consider the following two problems*

(I) *Find* $\boldsymbol{x} \in \mathbb{R}^m$ *such that:* $\quad \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \quad$ *and* $\quad \boldsymbol{x} \geq 0;$

(II) *Find* $\boldsymbol{\pi} \in \mathbb{R}^n$ *such that:* $\quad \boldsymbol{A}^T\boldsymbol{\pi} \leq \boldsymbol{0} \quad$ *and* $\quad \boldsymbol{b}^T\boldsymbol{\pi} > 0;$

then exactly only one of them has a solution.

Proof.

⇒ If (I) IS feasible the (II) IS NOT feasible:

Let (I) has a feasible solution, say $\boldsymbol{x} \geq \boldsymbol{0}$, then $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ so if there is a solution to (II), say $\boldsymbol{\pi}$, then $\boldsymbol{x}^T\boldsymbol{A}^T\boldsymbol{\pi} = \boldsymbol{b}^T\boldsymbol{\pi} > 0$. But then $\boldsymbol{A}^T\boldsymbol{\pi} > \boldsymbol{0}$ (since $\boldsymbol{x} \geq \boldsymbol{0}$), a contradiction. Hence (II) is infeasible.

Proof. (1/5).

$\Rightarrow$ If (I) IS NOT feasible then (II) IS feasible:

Let $C = \{\boldsymbol{z} \in \mathbb{R}^m \mid \boldsymbol{z} = \boldsymbol{A}\boldsymbol{x}, \boldsymbol{x} \geq \boldsymbol{0}\}$. If (I) is infeasible then $\boldsymbol{b} \notin C$. The set C is convex and closed (see next slides) so by the Separation Theorem there exists a real α and a vector $\boldsymbol{\pi}$ such that $\boldsymbol{b}^T\boldsymbol{\pi} > \alpha$ and $\boldsymbol{z}^T\boldsymbol{\pi} \leq \alpha$ for all $\boldsymbol{z} \in C$, that is,

$$\boldsymbol{x}^T\boldsymbol{A}^T\boldsymbol{\pi} \leq \alpha, \qquad \forall \boldsymbol{x} \geq \boldsymbol{0}$$

Since $\boldsymbol{0} \in C$ it follows that $\alpha \geq 0$, so $\boldsymbol{b}^T\boldsymbol{\pi} > 0$. If there exists an $\boldsymbol{z} \geq \boldsymbol{0}$ such that $\boldsymbol{z}^T\boldsymbol{A}^T\boldsymbol{\pi} > 0$ then

$$\lim_{\lambda \to \infty} (\lambda \boldsymbol{z}^T)\boldsymbol{A}^T\boldsymbol{\pi} = \infty$$

Therefore we must have $\boldsymbol{x}^T\boldsymbol{A}^T\boldsymbol{\pi} \leq 0$ for all $\boldsymbol{x} \geq \boldsymbol{0}$, and this holds if and only if $\boldsymbol{A}^T\boldsymbol{\pi} \leq \boldsymbol{0}$, which means that (II) is feasible.

Proof. (2/5).

The set C is convex:

Let $C = \{\boldsymbol{z} \in \mathbb{R}^m \mid \boldsymbol{z} = \boldsymbol{A}\boldsymbol{x}, \boldsymbol{x} \geq \boldsymbol{0}\}$. Let $\boldsymbol{z}_1$ and $\boldsymbol{z}_2 \in C$ then there exists $\boldsymbol{x}_1 \geq \boldsymbol{0}$ and $\boldsymbol{x}_2 \geq \boldsymbol{0}$ such that

$$\begin{aligned} \boldsymbol{z}_1 &= \boldsymbol{A}\boldsymbol{x}_1 \\ \boldsymbol{z}_2 &= \boldsymbol{A}\boldsymbol{x}_2. \end{aligned}$$

Moreover

$$\begin{aligned} \alpha \boldsymbol{z}_1 + (1-\alpha)\boldsymbol{z}_2 &= \boldsymbol{A}\big(\alpha \boldsymbol{x}_1 + (1-\alpha)\boldsymbol{x}_2\big), \\ \alpha \boldsymbol{x}_1 + (1-\alpha)\boldsymbol{x}_2 &\geq \boldsymbol{0}, \qquad \forall \alpha \in [0,1]. \end{aligned}$$

so that C is convex.

Proof. (3/5).

The set C is closed:

Let $\{\boldsymbol{z}_k\}$ a convergent sequence in C, i.e. $\lim_{k\to\infty} \boldsymbol{z}_k = \boldsymbol{z}$, for all k there exists $\boldsymbol{x}_k$ such that $\boldsymbol{z}_k = \boldsymbol{A}\boldsymbol{x}_k$ we choose $\boldsymbol{x}_k$ such that

$$\boldsymbol{z}_k = \boldsymbol{A}\boldsymbol{x}_k, \qquad \text{and} \qquad \boldsymbol{x}_k \perp \mathrm{Ker}(\boldsymbol{A})$$

if $\|\boldsymbol{x}_k\|$ is bounded $\|\boldsymbol{x}_k\|$ lie in a compact and thus there exists a subsequence such that

$$\lim_{j\to\infty} \boldsymbol{x}_{k_j} = \boldsymbol{x}, \qquad \boldsymbol{x} \geq \boldsymbol{0}.$$

and thus

$$\boldsymbol{z} = \lim_{j\to\infty} \boldsymbol{z}_{k_j} = \lim_{j\to\infty} \boldsymbol{A}\boldsymbol{x}_{k_j} = \boldsymbol{A} \lim_{j\to\infty} \boldsymbol{x}_{k_j} = \boldsymbol{A}\boldsymbol{x} \in C$$

so that C is closed.

Proof. (4/5).

if $\|\boldsymbol{x}_k\|$ is unbounded we have

$$\lim_{j\to\infty} \frac{\boldsymbol{z}_{k_j}}{\left\|\boldsymbol{x}_{k_j}\right\|} = \frac{\lim_{j\to\infty} \boldsymbol{z}_{k_j}}{\lim_{j\to\infty} \left\|\boldsymbol{x}_{k_j}\right\|} = \frac{\boldsymbol{z}}{\infty} = \boldsymbol{0}$$

we define the sequence $\boldsymbol{w}_j = \boldsymbol{x}_{k_j} / \left\|\boldsymbol{x}_{k_j}\right\|$ which is bounded and thus has a converging subsequence:

$$\lim_{i\to\infty} \boldsymbol{w}_{j_i} = \boldsymbol{w}, \qquad \|\boldsymbol{w}\| = 1, \quad \boldsymbol{w} \geq \boldsymbol{0}.$$

notice that

$$\boldsymbol{A}\boldsymbol{w} = \lim_{i\to\infty} \boldsymbol{A}\boldsymbol{w}_{j_i} = \lim_{i\to\infty} \frac{\boldsymbol{A}\boldsymbol{x}_{k_j}}{\left\|\boldsymbol{x}_{k_{j_i}}\right\|} = \lim_{i\to\infty} \frac{\boldsymbol{z}_{k_{j_i}}}{\left\|\boldsymbol{x}_{k_{j_i}}\right\|} = \boldsymbol{0}$$

and thus $\boldsymbol{w}$ is in the kernel of $\boldsymbol{A}$.

Proof. (5/5).

But for all $\boldsymbol{p} \in \mathrm{Ker}(\boldsymbol{A})$ we have

$$0 = \lim_{i\to\infty} \boldsymbol{p} \cdot \boldsymbol{w}_{j_i} = \boldsymbol{p} \cdot \lim_{i\to\infty} \boldsymbol{w}_{j_i} = \boldsymbol{p} \cdot \boldsymbol{w}$$

so that $\boldsymbol{w} \perp \mathrm{Ker}(\boldsymbol{A})$ and $\boldsymbol{w} \in \mathrm{Ker}(\boldsymbol{A})$ and thus $\boldsymbol{w} = \boldsymbol{0}$, a contradiction!. □

References

R. Tyrrell Rockafellar
Convex Analysis
Princeton University Press, 1996.

J. Farkas
Theorie der einfachen Ungleichungen
Journal für die reine und angewandte Mathematik, pp.1–27, **124**, 1902.

http://en.wikipedia.org/wiki/Multi-index_notation